BMJ Open

Developing an AI-assisted digital auscultation tool for automatic assessment of the severity of mitral regurgitation: protocol for a cross-sectional, non-interventional study

Li Zhang,[1] Zhenfeng Cheng,[1] Dongyang Xu,[2] Zhi Wang,[2,3] Shengsheng Cai,[2,4] Nan Hu,[5] Jianming Ma,[6] Xueqin Mei [7]

To cite: Zhang L, Cheng Z, Xu D, *et al*. Developing an AI-assisted digital auscultation tool for automatic assessment of the severity of mitral regurgitation: protocol for a cross-sectional, non-interventional study. *BMJ Open* 2024;**14**:e074288. doi:10.1136/bmjopen-2023-074288

► Prepublication history for this paper is available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2023-074288).

LZ and ZC contributed equally.

Received 11 April 2023
Accepted 11 March 2024

For numbered affiliations see end of article.

Correspondence to
Xueqin Mei;
meixueqin1970@163.com

ABSTRACT

Introduction Mitral regurgitation (MR) is the most common valvular heart disorder, with a morbidity rate of 2.5%. While echocardiography is commonly used in assessing MR, it has many limitations, especially for large-scale MR screening. Cardiac auscultation with electronic stethoscope and artificial intelligence (AI) can be a fast and economical modality for assessing MR severity. Our objectives are (1) to establish a deep neural network (DNN)-based cardiac auscultation method for assessing the severity of MR; and (2) to quantitatively measure the performance of the developed AI-based MR assessment method by virtual clinical trial.

Methods and analysis In a cross-sectional design, phonocardiogram will be recorded at the mitral valve auscultation area of outpatients. The enrolled patients will be checked by echocardiography to confirm the diagnosis of MR or no MR. Echocardiographic parameters will be used as gold standard to assess the severity of MR, classified into four levels: none, mild, moderate and severe. The study consists of two stages. First, an MR-related cardiac sound database will be created on which a DNN-based MR severity classifier will be trained. The automatic MR severity classifier will be integrated with the Smartho-D2 electronic stethoscope. Second, the performance of the developed smart device will be assessed in an independent clinical validation data set. Sensitivity, specificity, precision, accuracy and F1 score of the developed smart MR assessment device will be evaluated. Agreement on the performance of the smart device between cardiologist users and patient users will be inspected. The interpretability of the developed model will also be studied with statistical comparisons of occlusion map-guided variables among the four severity groups.

Ethics and dissemination The study protocol was approved by the Medical Ethics Committee of Huzhou Central Hospital, China (registration number: 202302009-01). Informed consent is required from all participants. Dissemination will be through conference presentations and peer-reviewed journals.

Trial registration number ChiCTR2300069496.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ An artificial intelligence-assisted automatic mitral regurgitation severity classifier will be developed and integrated with the Smartho-D2 electronic stethoscope.
⇒ An independent clinical validation data set will be established to enhance the reliability and generalisability of the experiment.
⇒ This study is non-interventional.
⇒ Simultaneously recorded ECG is not used as reference for phonocardiogram cycle segmentation.

INTRODUCTION

Mitral regurgitation (MR) is recognised as one type of commonly occurring valvular heart disorder with a high morbidity rate. An early study of the global burden of valvular heart disorder[1] reported that the morbidity rate of MR is as high as 2.5% among all populations and that this rate can increase substantially with increasing age. In the USA, more than two million adults have had MR and this number is predicted to double in 2030.[2] A large-scale, community, cross-sectional epidemiological study from Europe[3] revealed that the new diagnostic rate of moderate or severe MR is 2.3%, much larger than that of aortic stenosis (0.7%). A multicentre investigation study reported that MR is the most common valvular heart disorder in China.[4]

MR consists of two categories: primary/organic MR and secondary/functional MR. Caused by degenerative changes such as Barlow's disease or rheumatic heart diseases, organic MR is due to disruption of the valve leaflets, chordae tendineae and annulus. Compared with organic MR, functional MR (FMR) is more prevalent.[5] FMR is secondary to functional or structural abnormalities of

the heart, resulting in an imbalance between mitral valve tethering and closure force and hence poor coverage. Patients with ischaemic heart disease, dilated cardiomyopathy, heart failure with preserved ejection fraction or atrial fibrillation are apt to have FMR. Pharmacological therapy, cardiac resynchronisation therapy (CRT) or surgical/percutaneous interventions can be applied in the treatment of patients with MR[6] depending on the results of the MR assessment. It is very important that MR is detected and assessed in time as moderate or severe MR is highly associated with hospitalisation and long-term prognosis of left ventricular dysfunction and congestive heart failure.[7] However, in the USA, around 49% of patients with moderate or severe MR remain undiagnosed,[8] and this rate may be increased in developing countries. Therefore, a reliable, flexible and economical tool that can detect MR and assess its severity is essential.

Echocardiography is a key imaging technique for assessing the severity of MR. The severity level of MR is commonly categorised into three levels: mild, moderate and severe. Qualitative, semiquantitative and quantitative methods are used to assess the severity level of MR.[9] Quantitative methods have been well validated, while there is still no single echocardiographic parameter or optimal criteria to define the severity level of MR. In the recommendations for non-invasive evaluation of native valvular regurgitation from the American Society of Echocardiography,[10] effective regurgitant orifice area (EROA), regurgitant volume (RVol) and regurgitant fraction (RF) can be used to assess the severity level of MR. An EROA <0.2 cm^2, RVol <30 mL/beat or RF <30% is considered mild MR; 0.2 cm^2 ≤ EROA ≤ 0.39 cm^2, 30 mL/beat ≤ RVol ≤ 59 mL/beat or 30% ≤ RF ≤ 49% is considered moderate MR; and EROA ≥0.4 cm^2, RVol ≥60 mL/beat or RF ≥50% is considered severe MR. These parameters can be calculated by the proximal isovelocity surface area (PISA) method. Although echocardiographic parameters are commonly used to determine the severity level of MR, there are some drawbacks. First, most of the patients may not choose to go to hospital for assessment of their valvular heart disease due to lack or overconcentration of medical resources. Second, the assessment of echocardiography is highly operator-dependent as it is easily affected by transducer frequency, pulsed repetition frequency and inappropriate gain settings.[11] Lastly, the colour Doppler device is cumbersome and complex to use and hence not suitable for long-term monitoring of MR.

Cardiac auscultation is an important tool that is often used during first-time cardiac examinations. Although cardiac auscultation has the advantage of quick examination, it requires plenty of experience and skills. Their perceptive acoustic frequency ranges are different for different users. In comparison with traditional auscultation, with the aid of emerging electronic stethoscopes,[12] smart auscultation manipulates digitalised phonocardiogram (PCG) with digital signal processing and pattern recognition techniques. In abnormal PCG recordings, apart from the first heart sound (S1) and the second heart sound (S2), other acoustic components such as the third heart sound (S3), the fourth heart sound (S4), murmurs, gallops, clicks or opening snaps may appear. Features of the PCG signal corresponding to specific heart disease can be extracted and used for PCG classification. Dwivedi *et al*[13] conducted a systematic review of automatic PCG analysis and classification, where conventional steps in this task were concluded as preprocessing (denoising), segmentation, feature extraction and classification. The review has now been updated,[14] where deep learning algorithms for heart sound classification in recent years have been supplemented. For cardiac auscultation denoising, various digital signal processing methods[15] have been implemented, and the current trend is to combine these conventional methods with deep neural networks (DNNs).[16] The necessity of PCG cycle segmentation is controversial, as in some studies reasonable PCG classification performance has also been achieved without segmentation.[17] Feature extraction is important for PCG signal classification, where handcrafted features[18] including time-frequency spectrum, Mel frequency cepstrum coefficients (MFCCs) and discrete wavelet transform coefficients have been considered, and in recent years automatic feature extraction from the PCG spectra or waveform by DNNs has been prevalent.[19] For handcrafted feature extraction, it is always followed by traditional machine learning methods, including support vector machine, random forest, Gaussian mixture model, etc. In deep learning methods such as convolutional neural network (CNN) or recurrent neural network, the classifiers are typically trained simultaneously with feature extraction step.

Although various PCG signal classification methods are coming forth, most of them are confined to distinguishing normal PCG recordings from abnormal ones, restricted by the task set-up in two commonly used public databases: the Physionet heart sound challenge 2016 database[20] and the PASCAL classifying heart sounds challenge 2011 database.[21] Recently, a database for detection of heart valve disorders was shared in the GitHub repository.[22] This database includes 1000 PCG segments, each one lasting for around 2 s, truncated from several PCG recording sources in books and websites. The segments were divided evenly into five categories: normal, aortic stenosis, MR, mitral stenosis and mitral valve prolapse. Several works were carried out in handling this five-category classification task.[23–29] Time-frequency magnitude and phase features were used,[23] and graph-based feature was developed,[24] with traditional machine learning methods used for classification. Avanzato and Beritelli[25] discovered that with raw data as the input to 1-dimensional CNN (1D-CNN), reasonable classification performance can still be achieved. Karhade *et al*[26] used both time domain polynomial chirplet transform and frequency domain polynomial chirplet transform for CNN-based classification. Convolutional and recurrent neural networks were developed by Alkhodari and Fraiwan[27] and attention mechanism was introduced by

Chowdhury *et al*[28] to address this issue. Recursive feature elimination algorithm was applied by Arslan[29] to select the most distinctive deep features for heart valve disorder classification. Although the above-mentioned methods may show potential values in MR detection, there are no studies that aim to assess the severity of MR due to the lack of data as well as methods. In spite of this, the existing studies on PCG-based assessment of severity of other heart valve disorders, such as aortic stenosis[30] and tricuspid regurgitation,[31] can be used for reference.

The Huzhou Institute of Zhejiang University and Melodicare developed an electronic stethoscope device named Smartho-D2, which has been certificated by Food and Drug Administration (FDA) 510k, National Medical Products Administration (NMPA) and Conformité Européenne (CE). Smartho-D2 has two sound pick-up channels, one for auscultation and another for background noise collection. Our developed two-stage noise cancellation algorithm[16] has been applied to this device to eliminate background noise contamination. Smartho-D2 is portable, Bluetooth-connected and has built-in software supporting PCG waveform displaying, data storage and download from local and the cloud, online remote auscultation, etc. Secondary software development can be easily carried out in Smartho-D2 to add more machine learning-based heart disease assessment functionalities. In this study, Smartho-D2 will be used for PCG recording from patients with MR. With the recorded PCG signals labelled according to echocardiographic examination, an automatic MR severity assessment algorithm will be developed and integrated with Smartho-D2. The performance of the developed smart stethoscope in automatically assessing the severity of MR will be evaluated via clinical experiments.

Objectives and hypotheses

The first objective of this study is to establish a DNN-based cardiac auscultation method that can automatically assess the severity of MR. Based on the findings of previous studies,[23–31] the hypothesis in this study is that putting PCG signals into a trained DNN can lead to automatic classification of the severity of MR, with high sensitivity and specificity. The second objective is to quantitatively measure the performance of the developed artificial intelligence (AI)-based MR assessment method when the AI-aided electronic stethoscope Smartho-D2 is used by cardiologists as well as patients. It is hypothesised that the developed smart stethoscope will yield high classification performance metrics in automatically assessing the severity of MR and that the agreement between cardiologists and patients on the performance of the device will be high.

METHODS AND ANALYSIS

Sample selection and patient recruitment

This study and the related experiments will be carried out at the Department of Cardiology, Huzhou Central Hospital, which is located north of Zhejiang Province in China, providing healthcare to about four million people. Each year, about 50 000 outpatients come to seek help from a cardiologist and over 50% of them suffer from heart valve diseases. By a cross-sectional design, adult patients diagnosed with MR or no MR will be considered for enrolment into the experiments. The following are the MR patient inclusion criteria: (1) the patient has been diagnosed to have MR by echocardiographic examination; (2) the patient did not take any medicine that can influence cardiac function in 24 hours; and (3) the patient did not have a history of heart valve surgery. If the quality of PCG signal recorded from a participant is too low to be used, it will be excluded. There will be no limitations to gender and age of patients with MR. Randomisation will not be used for patients with MR. Patients with no MR will be matched to patients with MR by propensity score. The study will be conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Written informed consent will be obtained from the patients before the data recording experiments.

Study design

There will be two stages in implementing the study. The first stage aims at building an automatic MR severity classifier with PCG signal as input. A PCG database for training the MR severity classifier will be built. The database will include PCG recordings collected from recruited patients mentioned above and the data labels will be the four levels of MR severity: none, mild, moderate and severe. Echocardiography will be used as gold standard to label the severity of MR. A DNN will be built to form an automatic MR severity classifier and trained based on our MR-related PCG database. The trained MR severity classifier will be integrated with the Smartho-D2 electronic stethoscope. In the second stage, the performance of AI-aided electronic stethoscope Smartho-D2 for automatic MR severity assessment will be evaluated in real clinical setting. In this stage, the gold standard echocardiographic examination will also be performed to label the severity of MR. As bias may occur when a smart stethoscope is used by different types of operators, agreement on the results of the automatic MR severity assessment between a cardiologist user and a patient user will be evaluated. The overall flow of the study is shown in figure 1.

For each patient, a cardiologist will collect one PCG recording at the mitral valve auscultation area for at least 10 s using the electronic stethoscope Smartho-D2. The mitral valve auscultation area is at the cardiac apex, located in the fifth intercostal space on the midclavicular line. Smartho-D2 has a wide frequency response, ranging from 20 Hz to 20 000 Hz. Evaluated by testing acoustic sources, it was found that a 13 dB gain was achieved with 100–500 Hz, 13–21 dB gain was achieved with 500–600 Hz and 21 dB gain was maintained with 600–1000 Hz. The sampling rate is 8000 Hz. As primary heart sound components and murmurs typically have frequency ranges

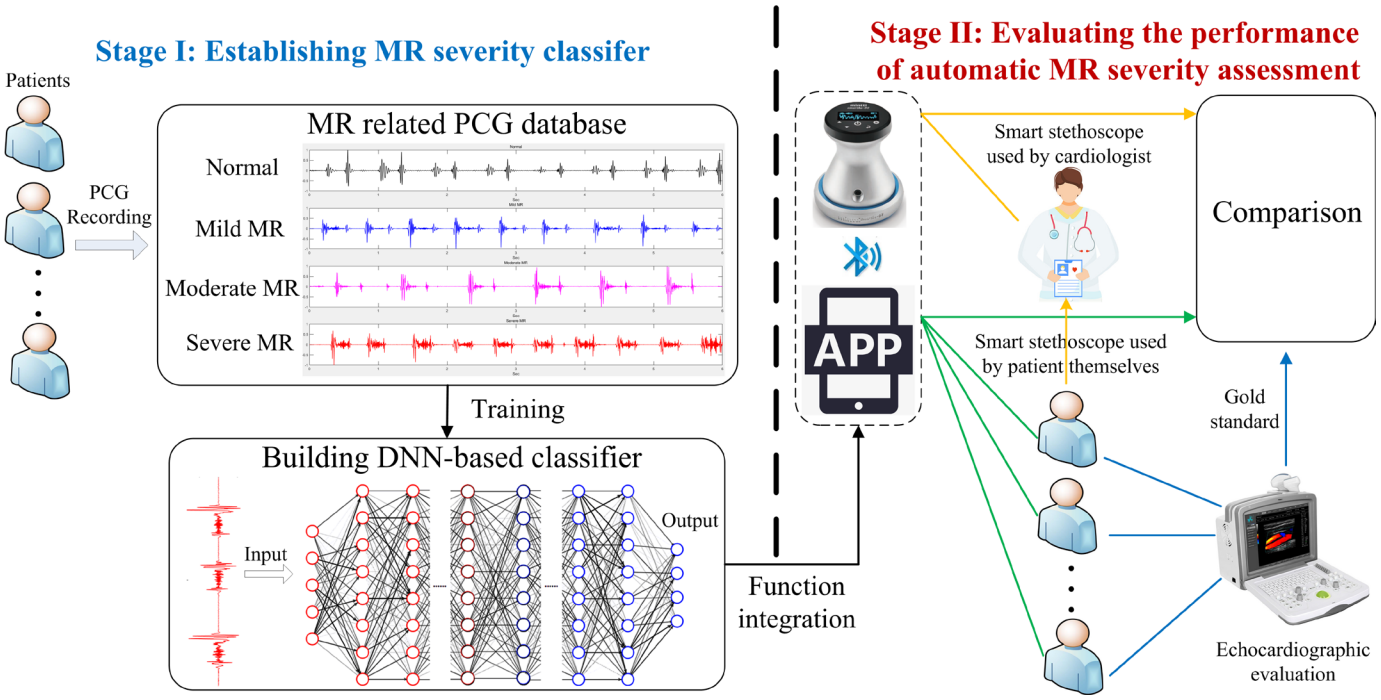

Figure 1 Study flow chart. DNN, deep neural network; MR, mitral regurgitation; PCG, phonocardiogram.

below 250 Hz and 600 Hz, respectively, the frequency response characteristics and the sampling rate can meet the high-quality requirement of PCG recording. Smartho-D2 is Bluetooth-connected, and the Smartho app is used to monitor and control the data recording. The recorded data can be labelled and stored in a local mobile device and can also be transmitted to a password-protected database server located at Huzhou Central Hospital. The storage format of each PCG recording is the uncompressed '.pcm', and no encoding or compression algorithms were used when sending data to the cloud server. An outline of the PCG data recording and storage in this study is shown in figure 2. The recruited patients accept a routine echocardiographic examination to label the severity level of their MR. According to the recommendations for non-invasive evaluation of native valvular regurgitation from the American Society of Echocardiography,[10] EROA, RVol and RF calculated by the PISA method will be used for labelling. Mild MR is considered when at least two of the following conditions are satisfied: 0 cm^2 < EROA < 0.2 cm^2, 0 mL/beat < RVol < 30 mL/beat, 0 < RF < 30%. Moderate MR is considered when at least two of the following conditions are satisfied: 0.2 cm^2 ≤ EROA ≤ 0.39 cm^2, 30 mL/beat ≤ RVol ≤ 59 mL/beat, 30% ≤ RF ≤ 49%. Severe MR is considered when at least two of the following conditions are satisfied: EROA ≥0.4 cm^2, RVol ≥60 mL/beat or RF ≥50%. Otherwise no MR will be considered. Examinations beyond the routine ones in this study are cost-free. After protected health information is deidentified, the data can be used by non-medical members for further processing and analysis.

The automatic MR severity classifier to be built will further be integrated with the Smartho-D2 electronic stethoscope. For practical usage, it will consist of adaptive

Figure 2 Outline of the phonocardiogram data recording and storage.

noise cancellation, usable data extraction and the trained DNN. Smartho-D2 has two sound pick-up channels, and a two-stage noise cancellation algorithm[16] has been applied to this device. As reasonable PCG classification performance has been achieved without cardiac cycle segmentation in some previous work,[17] in this study we will only divide the PCG recording into segments with sliding windows. A 2 s window will be used, with a sliding stride of 1 s. In real applications, PCG recording may be contaminated by inappropriate operations such as pressure with excessive force, clothing friction and auscultation leaving the body surface, especially when the stethoscope is used by the patient themselves. In this study, before being put into the trained DNN, each PCG segment will be evaluated by a previously developed usable data extraction module.[16] With a simple CNN structure, it judges a PCG segment as usable or unusable, where an unusable PCG segment implies that its quality is too low to be used for MR severity assessment. Only usable PCG segments will be put into the DNN trained with the established data set.

It has been recognised that the frequency range of pathological murmur is often wider than that of primary heart sounds, and therefore time-frequency spectrum or MFCCs are usually chosen as the input features to the DNN.[18] However, frequency domain features may only be part of the useful features that classify the severity level of MR. Benefited by the power of efficiency in automatic feature extraction, the DNN built in this study is expected to extract and use features from data as much as possible for the classification task, with PCG segment as the input. In this study, a clique block-based DNN[32 33] will be established, which is believed to take both feature reuse and lightweight features into account. A feasible DNN structure is given at the top of figure 3, where features are automatically extracted by the clique blocks in different levels and concatenated for final classification. 1D convolutions

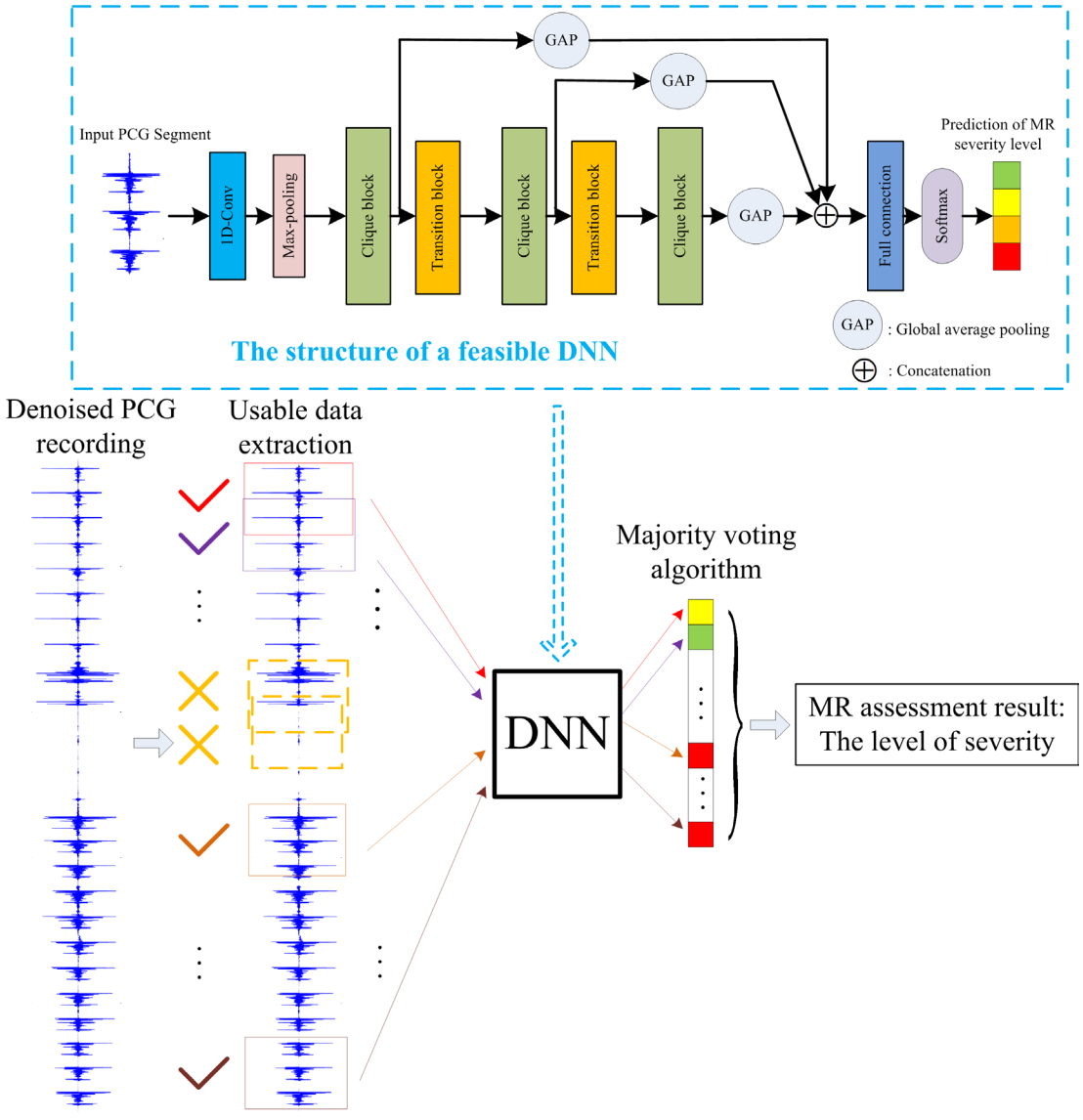

Figure 3 Flow chart of PCG data processing and structure of a feasible DNN to be used in the automatic MR severity classifier. DNN, deep neural network; MR, mitral regurgitation; PCG, phonocardiogram.

are used in particular to tackle the processing of 1D signals or 1D features. The transition block involves channel-wise attention mechanism and uses convolution and pooling to achieve dynamic feature calibration. As the MR severity classifier involving a trained DNN will be integrated with a mobile device, the neural architecture search[34] will be used to refine the performance and model size of the DNN. For each usable PCG segment, the trained DNN will provide a four-class classification result as output. For comprehensive utilization of the classification results from all usable PCG segments, a majority voting strategy will be applied to give the final decision on the level of MR severity. The flow chart of the PCG data processing is shown in figure 3.

Interventions

There will be no intervention to be implemented to the patients in this study. The subjects in this study are patients diagnosed with MR or no MR by cardiologists, and our objective relies on exploiting the performance of automatic assessment of the severity level of MR with PCG signals. The study will not impact the diagnosis or further clinical care.

Sample size calculation

In the first stage of the study, the main task is to collect PCG data and train the automatic MR severity classifier. The established database will be divided into three pieces equally, and a threefold cross-validation will be performed. In each round of model training, two pieces are used as the training set and one piece is used as the testing set, and finally the model with median classification metrics among the three trained ones will be adopted. Let a one-sample proportion test, the z-test with continuity correction, be used to assess whether the accuracy of the trained classifier $P1$ is significantly larger than a hypothesised value $P0$. To detect $P1$=0.95 compared with $P0$=0.90, with one-sided type I error of 0.05 and power of 0.80 calculated by binomial enumeration, the required sample size in a testing set is 179. Hence, the total required sample size in the first stage of the study is 179×3=537.

In the second stage of the study, all collected PCG data will be used as the testing set. In this stage, the PCG as well as the results of the automatic assessment by the MR severity classifier integrated with the electronic stethoscope will be recorded. This independent clinical validation data set aims at enhancing the reliability and generalisability of the experiment. Compared with the first stage, the required sample size in the second stage seems to be 179. Cohen's kappa will be used to evaluate the agreement on the results of the automatic MR severity assessment between a cardiologist and a patient. When the value of kappa under the null hypothesis is $\kappa 0$=0.5, the one under the alternative hypothesis is $\kappa 1$=0.6 and the proportions of subjects assigned to four categories are 0.3:0.3:0.2:0.2, for a one-sided alternative hypothesis test of $\kappa 1 > \kappa 0$ and with type I error of 0.05 and power of 0.80, the required sample size is calculated as 286. As 286 >179, the required sample size in the second stage of the study should be 286.

Finally, the total required sample size in this study will be 537+286=823. PASS (Power Analysis and Sample Size) V.2021 software is used to calculate the size of the sample to be collected. The schedule of data collection is outlined in table 1.

Table 1 Schedule of data collection

Research stage	Class of patients	Sample size
Stage 1: building the training data set for the classifier (15 March 2023–15 September 2024)	No MR	161
	Mild MR	161
	Moderate MR	107
	Severe MR	108
	Total	537
Stage 2: independent clinical validation (16 September 2024–15 March 2025)	No MR	86
	Mild MR	86
	Moderate MR	57
	Severe MR	57
	Total	286
Total		823

MR, mitral regurgitation.

Statistical analysis

The aim of the first stage is to build an automatic MR severity classifier. Blinding will be considered in the testing sets. Classification metrics, including sensitivity, specificity, precision, accuracy and F1 score, will be calculated and assessed. As a multicategory classification task is involved in this study, macro-average, micro-average and weighted average metrics will be calculated. Macro-average means the direct average of the binary classification metrics across classes. The micro-average metrics are calculated using aggregate outcomes across all classes. The weighted average metrics are given by the weighted average of the binary classification metrics across classes, with proportions of classes as weights. Their one-versus-rest receiver operating characteristic curves and the corresponding area under the curve with CI will also be given.

The above-mentioned classification metrics will also be assessed in the second stage, where the performance of the established MR severity assessment system will be evaluated with real clinical usage. In this stage, to measure the bias when a smart stethoscope is used by different types of users, the agreement on the results of the MR severity assessment between a cardiologist user and a patient user will be evaluated using the metrics of Cohen's kappa. A kappa score >0.8 will be deemed a good agreement.

The interpretability of the developed classification model will also be considered. Inspired by a robust interpretable deep learning-based PCG classifier study,[17] occlusion maps will be employed in this study to show which area (S1, systole, S2 or diastole) in each PCG cycle

is automatically emphasised by the model for MR severity classification. After the emphasised areas are determined, three classes of computerised feature variables, including temporal domain features, frequency domain features and chaotic features, will be calculated from these areas for statistical comparison among the four severity groups (none, mild, moderate and severe). One-way analysis of variance will be used first, and then Student-Newman-Keuls q-test (SNK-q) will be used for pairwise comparison. P<0.5 will be considered statistically significant. The above deep learning-guided data mining for significant variables is expected to show the interpretability of the developed model.

Data collection and analysis plan

The data collection started on 15 March 2023 and is planned to last for 2 years. The first stage of the study is planned to be finished on 15 September 2024. Considering about 50 000 outpatient visits in 1 year, we believe that a 1-year first-stage data collection plan will easily meet the sample size requirement for training, even with a conservative MR morbidity rate of 2.5% for these outpatients. The Smartho app, deployed in a password-protected Android pad, is used to monitor and control the PCG data recording. Echocardiographic examination reports given by a skilled expert with above 10-year experience will be used in data labelling. The labelled data will be transferred from local device to a password-protected database server located at Huzhou Central Hospital, and anonymised data files can be further processed and analysed by non-medical members. Data analysis to establish the classification model and integrate with the device will be started on 15 March 2024. With the growth of recorded data size, the weights in the classification model will be updated through training, and the classification performance is expected to be refined.

Patient and public involvement

None.

Ethics and dissemination

Ethical approval was obtained from the Medical Ethics Committee of Huzhou Central Hospital, China (registration number: 202302009-01). This study was registered with the Chinese Clinical Trial Registry (ChiCTR2300069496). Before PCG recording, written informed consent will be obtained from all patients. Access to the collected data will be restricted to cardiologists from the research team treating the participants, and anonymised data files will be shared to non-medical members of the research team. Dissemination will be through conference presentations and peer-reviewed journals.

DISCUSSION

Although it has been realised a long time ago that quantitatively analysing PCG signal may provide a convenient and fast way to assess heart valve diseases, there is no PCG-based method or device that automatically assesses the severity of MR, the most common valvular heart disorder. In recent years, using a publicly shared heart valve disorder PCG database,[22] many studies[23–29] have been published that tried to distinguish MR from normal and the other three valvular heart disorders. From these existing studies, MR-related PCG signals show specific pattern that can be employed in MR detection. Although

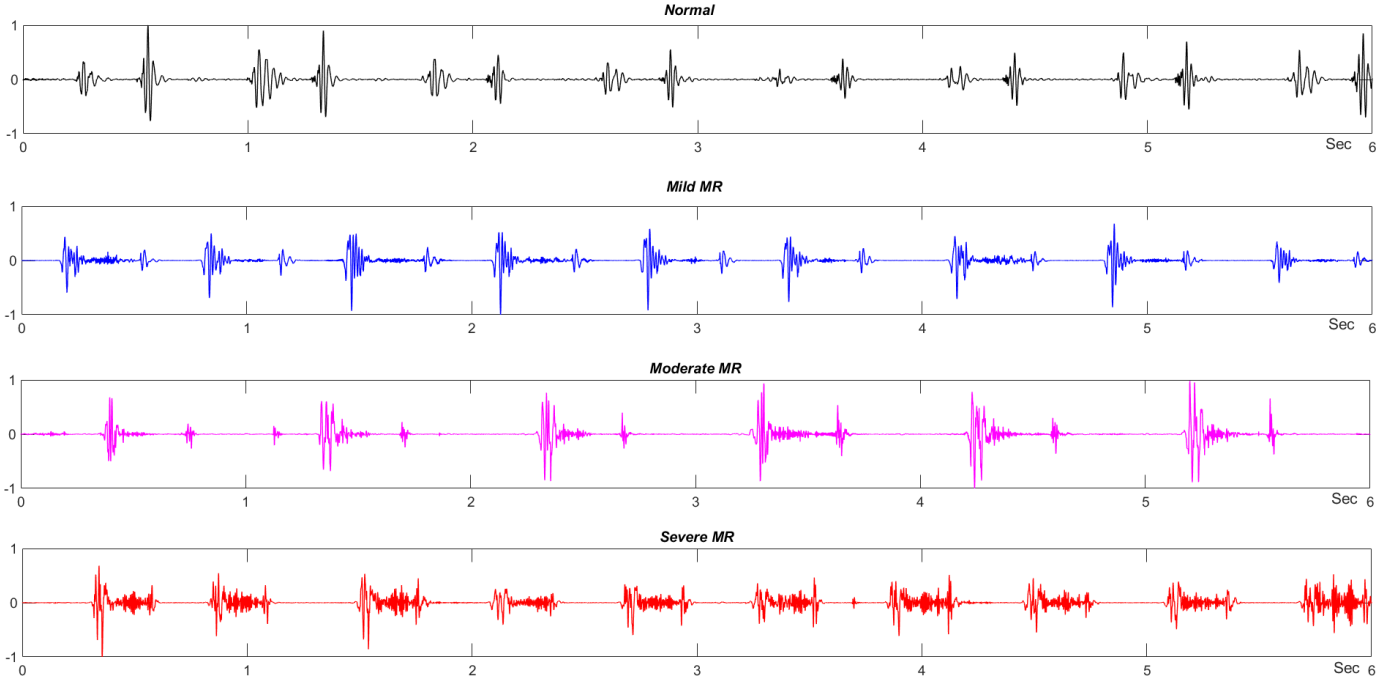

Figure 4 Typical PCG cases corresponding to different levels of MR severity collected in this study (data collection began on 15 March 2023). MR, mitral regurgitation; PCG, phonocardiogram.

assessing the severity of MR using PCG signal has not been studied in the past, similar methodology has been applied to automatically grade the severity level of aortic stenosis[30] and tricuspid regurgitation.[31] By studying the PCG signals and echocardiographic parameters of 41 patients with aortic stenosis,[30] it was found that there was a good correlation between the duration of spectra above 300 Hz and the Doppler-derived peak pressure gradient. Rujoie *et al*[31] studied the PCG signals of 22 subjects (divided into three categories: none, mild and moderate to severe tricuspid regurgitation) using MFCC and wavelet transform methods for feature extraction, genetic algorithms for feature selection and K-Nearest-Neighbor (KNN) to implement the triple-classification task. The data recording has been started recently, and figure 4 shows the typical recorded PCG cases corresponding to the different levels of MR severity. It seems that even the waveforms may reflect some information for classification, for example, the pattern of systolic murmur. Different from research based on handcrafted features and small-sized database, our study considers larger-sized data collection and automatic feature extraction via DNN.

The application of deep learning in the analysis of PCG signals has become promising in recent years. The existing studies in this field, discriminating among five-class valvular heart disorders[22] for instance, involved various input features or network structures. To mimic the processing of images, time-frequency images are usually derived as input features to 2D CNNs. Karhade *et al*[26] used a two-layered 2D CNN, and Chowdhury *et al*[28] introduced attention mechanism to the 2D CNN. Arslan[29] used continuous wavelet transform to generate scalogram images and employed multilayered extreme learning machine and several pretrained DNNs, including VGG16, ResNet50 and MobileNetV2, to extract deep features. Handcrafted features may cause loss of information for classification, and therefore several studies tried to use 1D convolutions to extract features from a raw data segment. Avanzato and Beritelli[25] built a five-layered 1D CNN[25] to extract features from raw PCG data, while Alkhodari and Fraiwan[27] used a three-layered 1D CNN cascaded by a Bi-directional Long Short-Term Memory (BiLSTM). In this study, we consider directly processing the noise-cancelled PCG segment with a clique block-based DNN, where clique blocks and channel-wise attention are employed. The clique block has shown advantages in feature extraction and reuse, compared with other prevalent blocks such as residual blocks or dense blocks. Apart from the feature extraction issue, another major drawback of the existing deep learning-based studies is that the developed 'black boxes' may lead to ethically problematic outcomes.[35] Inspired by the pioneer research on the interpretability of the PCG classification model,[17] we use the trained deep learning model to guide the way to the discovery of areas in a PCG cycle that are important for classification. The interpretability will be further enhanced by statistically comparing computerised variables derived from these areas among groups.

The MR severity level classifier will be integrated with the Smartho-D2, an electronic stethoscope device certified by the FDA 510k, NMPA and CE. In recent years, some AI-aided auscultation products have emerged for cardiac healthcare. In 2019, the eMurmur Heart AI received FDA clearance and CE marking for detecting the presence of a murmur on PCG recording and determining if it is likely innocent or pathological. Evaluation of 603 outpatient visits[36] showed that the eMurmur had 93% (CI 90% to 95%) sensitivity and 81% (CI 85% to 91%) specificity in the detection of pathological cases. In 12 July 2022, the FDA cleared the Eko Murmur Analysis Software (EMAS) for detecting and characterising murmurs found in adult and paediatric patients. The EMAS is reported to identify structural murmurs with a sensitivity of 90.2% and specificity of 90.6% among adults and has an overall sensitivity of 85.6% and specificity of 84.4%.[37] These existing smart cardiac auscultation applications showed the feasibility of developing valvular heart disorder assessment methods on electronic stethoscope devices.

A limitation of this study is that only PCG will be used, without simultaneously recorded ECG. In many studies, ECG was used as reference for PCG cycle segmentation.[38] However, in recent years, some studies have shown that segmentation may not be necessary when deep learning is used for cardiac sound classification. By using Shapley values and occlusion maps, Dissanayake *et al*[17] showed that deep learning methods can automatically emphasise the components on PCG data that make significant contribution to the classification. In our study, we consider classifying the PCG signals without PCG cycle segmentation. The collected large-sized data will endow the established DNN-based MR severity level classifier with the ability to automatically select useful PCG components. A multicentre study involving more subjects in a virtual clinical trial will be among our future work to refine and evaluate the performance of the developed method and smart device.

Author affiliations
[1]Department of Cardiology, Huzhou Central Hospital, The Fifth School of Clinical Medicine of Zhejiang Chinese Medical University, The Affiliated Central Hospital of Huzhou University, Huzhou, Zhejiang, China
[2]Center for Intelligent Acoustics and Signal Processing, Huzhou Institute of Zhejiang University, Huzhou, Zhejiang, China
[3]State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou, Zhejiang, China
[4]Suzhou Melodicare Medical Technology Co., Ltd, Suzhou, Jiangsu, China
[5]School of Electronics and Information Engineering, Soochow University, Suzhou, Jiangsu, China
[6]Administration Office, Huzhou Central Hospital, The Fifth School of Clinical Medicine of Zhejiang Chinese Medical University, The Affiliated Central Hospital of Huzhou University, Huzhou, Zhejiang, China
[7]Department of Medical Engineering, Huzhou Central Hospital, The Fifth School of Clinical Medicine of Zhejiang Chinese Medical University, The Affiliated Central Hospital of Huzhou University, Huzhou, Zhejiang, China

Contributors LZ and ZC conceptualised and designed the study and wrote the original draft. DX, ZW and NH will analyse the recorded data using digital signal processing algorithms and design the DNN-based classifier. SC will be responsible for the technical support for the device. JM will be responsible for project

administration. XM provides mentorship to this study, edited and revised the draft, and will be responsible for project administration. All authors critically reviewed and approved the final manuscript before submission.

Funding This work was supported by the Huzhou Science and Technology Project (grant 2022GZ46) and the Cardiovascular Discipline Group Funds of Huzhou (grant XKQ-HT-202102A).

Competing interests SC is an employee and owner of Suzhou Melodicare Medical Technology, which developed the Smartho-D2 device.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD

Xueqin Mei http://orcid.org/0009-0000-1140-8560

REFERENCES

1 Nkomo VT, Gardin JM, Skelton TN, *et al*. Burden of valvular heart diseases: a population-based study. *Lancet* 2006;368:1005–11.
2 de Marchena E, Badiye A, Robalino G, *et al*. Respective prevalence of the different carpentier classes of mitral regurgitation: a stepping stone for future therapeutic research and development. *J Card Surg* 2011;26:385–92.
3 d'Arcy JL, Coffey S, Loudon MA, *et al*. Large-scale community echocardiographic screening reveals a major burden of undiagnosed valvular heart disease in older people: the Oxvalve population cohort study. *Eur Heart J* 2016;37:3515–22.
4 Zhang B, Xu H, Zhang H, *et al*. Prognostic value of N-terminal pro-B-type natriuretic peptide in elderly patients with valvular heart disease. *J Am Coll Cardiol* 2020;75:1659–72.
5 Dziadzko V, Dziadzko M, Medina-Inojosa JR, *et al*. Causes and mechanisms of isolated mitral regurgitation in the community: clinical context and outcome. *Eur Heart J* 2019;40:2194–202.
6 Otto CM, Nishimura RA, Bonow RO, *et al*. 2020 ACC/AHA guideline for the management of patients with valvular heart disease: executive summary: a report of the American college of cardiology American Heart Association joint committee on clinical practice guidelines. *Circulation* 2021;143:e35–71.
7 Sannino A, Smith RL, Schiattarella GG, *et al*. Survival and cardiovascular outcomes of patients with secondary mitral regurgitation: a systematic review and meta-analysis. *JAMA Cardiol* 2017;2:1130–9.
8 Head SJ, van Leeuwen WJ, Van Mieghem NM, *et al*. Surgical or transcatheter mitral valve intervention: complex disease requires complex decisions. *EuroIntervention* 2014;9:1133–5.
6 Otto CM, Nishimura RA, Bonow RO, *et al*. 2020 ACC/AHA guideline for the management of patients with valvular heart disease: executive summary: a report of the American college of cardiology American heart Association joint committee on clinical practice guidelines. *Circulation* 2021;143:e35–71.
10 Zoghbi WA, Adams D, Bonow RO, *et al*. Recommendations for noninvasive evaluation of native valvular regurgitation: a report from the American society of echocardiography developed in collaboration with the society for cardiovascular magnetic resonance. *J Am Soc Echocardiogr* 2017;30:303–71.
11 Stevenson JG. Two-dimensional color Doppler estimation of atrioventricular valve regurgitation: important effects on instrument gain setting, pulse repetition frequency and carrier frequency. *J Am Soc Echocardiogr* 1989;2:1–10.
12 Bank I, Vliegen HW, Bruschke AVG. The 200th anniversary of the stethoscope: can this low-tech device survive in the high-tech 21st century? *Eur Heart J* 2016;37:3536–43.
13 Dwivedi AK, Imtiaz SA, Rodriguez-Villegas E. Algorithms for automatic analysis and classification of heart sounds - A systematic review. *IEEE Access* 2019;7:8316–45.
14 Chen W, Sun Q, Chen X, *et al*. Deep learning methods for heart sounds classification: a systematic review. *Entropy (Basel)* 2021;23:667.
15 Mondal A, Saxena I, Tang H, *et al*. A noise reduction technique based on nonlinear kernel function for heart sound analysis. *IEEE J Biomed Health Inform* 2018;22:775–84.
16 Yang C, Dai N, Wang Z, *et al*. Cardiopulmonary auscultation enhancement with a two-stage noise cancellation approach. *Biomed Signal Process Control* 2023;79:104175.
17 Dissanayake T, Fernando T, Denman S, *et al*. A robust interpretable deep learning classifier for heart anomaly detection without segmentation. *IEEE J Biomed Health Inform* 2020;25:2162–71.
18 Xu W, Yu K, Ye J, *et al*. Automatic pediatric congenital heart disease classification based on heart sound signal. *Artif Intell Med* 2022;126:102257.
19 Li S, Li F, Tang S, *et al*. Heart sounds classification based on feature fusion using lightweight neural networks. *IEEE Trans Instrum Meas* 2021;70:1–9.
20 Clifford GD, Liu C, Moody B, *et al*. Classification of normal/abnormal heart sound recordings: the Physionet/computing in cardiology challenge 2016. Data from: PhysioNet Repository, version March 1, 2016. Proc Comput Cardiology Conf (CinC); 2016:609–12 Available: https://archive.physionet.org/challenge/2016
21 Bentley P, Nordehn G, Coimbra M, *et al*. Data from: the PASCAL classifying heart sounds challenge, version November 1, 2011. 2011. Available: http://www.peterjbentley.com/heartchallenge
22 Yaseen -, Son G-Y, Kwon S. Classification of heart sound signal using multiple features. *Appl Sci* 2018;8:2344.
23 Ghosh SK, Tripathy RK, Ponnalagu RN, *et al*. Automated detection of heart valve disorders from the PCG signal using time-frequency magnitude and phase features. *IEEE Sens Lett* 2019;3:1–4.
24 Tuncer T, Dogan S, Tan RS, *et al*. Application of Petersen graph pattern technique for automated detection of heart valve diseases with PCG signals. *Information Sciences* 2021;565:91–104.
25 Avanzato R, Beritelli F. Heart sound multiclass analysis based on raw data and convolutional neural network. *IEEE Sens Lett* 2020;4:1–4.
26 Karhade J, Dash S, Ghosh SK, *et al*. Time–frequency-domain deep learning framework for the automated detection of heart valve disorders using PCG signals. *IEEE Trans Instrum Meas* 2022;71:1–11.
27 Alkhodari M, Fraiwan L. Convolutional and recurrent neural networks for the detection of valvular heart diseases in Phonocardiogram recordings. *Comput Methods Programs Biomed* 2021;200:105940.
28 Chowdhury S, Morshed M, Fattah SA. Spectrocardionet: an attention-based deep learning network using triple-Spectrograms of PCG signal for heart valve disease detection. *IEEE Sensors J* 2022;22:22799–807.
29 Arslan Ö. Automated detection of heart valve disorders with time-frequency and deep features on PCG signals. *Biomed Signal Process Control* 2022;78:103929.
30 Kim D, Tavel ME. Assessment of severity of aortic stenosis through time-frequency analysis of murmur. *Chest* 2003;124:1638–44.
31 Rujoie A, Fallah A, Rashidi S, *et al*. Classification and evaluation of the severity of Tricuspid regurgitation using Phonocardiogram. *Biomed Signal Process Control* 2020;57:101688.
32 Yang Y, Zhong Z, Shen T, *et al*. Convolutional neural networks with alternately updated clique. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2018:2413–22
33 Xiao B, Xu Y, Bi X, *et al*. Follow the sound of children's heart: a deep-learning-based computer-aided pediatric CHDs diagnosis system. *IEEE Internet Things J* 2020;7:1994–2004.
34 Zoph B, Le QV. Neural architecture search with reinforcement learning. Proc Int Conf Learn Representations 2017.
35 Char DS, Shah NH, Magnus D. Implementing machine learning in health care - addressing ethical challenges. *N Engl J Med* 2018;378:981–3.
36 Thompson WR, Reinisch AJ, Unterberger MJ, *et al*. Artificial intelligence-assisted Auscultation of heart murmurs: validation by virtual clinical trial. *Pediatr Cardiol* 2019;40:623–9.
37 Eko murmur analysis software user manual. n.d. Available: https://www.ekohealth.com
38 Zhang H, Wang X, Liu C, *et al*. Discrimination of patients with varying degrees of coronary artery stenosis by ECG and PCG signals based on entropy. *Entropy (Basel)* 2021;23:823.